DOCUMENT RESUME

ED 370 999 TM 021 610

AUTHOR Thompson, Bruce
TITLE The Revised Program Evaluation Standards and Their Correlation with the Evaluation Use Literature.
PUB DATE Apr 94
NOTE 41p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 4-8, 1993).
PUB TYPE Information Analyses (070) -- Reports - Descriptive (141) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Evaluation Methods; *Evaluation Utilization; *Evaluators; Literature Reviews; *Program Evaluation; *Responsibility; *Standards
IDENTIFIERS *Joint Committee on Standards for Educ Evaluation

ABSTRACT

Various strategies for optimizing the use of program evaluation information are recommended, based on a thorough evaluation of the literature. The recently revised "Program Evaluation Standards" (Joint Committee, 1994) are then reviewed point by point as regards sensitivity to these recommendations. It appears that the new standards generally incorporate what the literature suggests should be regarded as "best practice." The revised standards place the onus of considerable responsibility for promoting use on the shoulders of the evaluators themselves. (Contains 153 references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

evaluse.wp1 4/1/94

ED 370 999

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
BRUCE THOMPSON
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

The Revised Program Evaluation Standards and Their Correlation with the Evaluation Use Literature

Bruce Thompson

Texas A&M University 77843-4225
and
Baylor College of Medicine

Paper presented at the annual meeting (session #44.27) of the American Educational Research Association, New Orleans, LA, April 7, 1994. Portions of this paper were adapted from the previous report of Thompson and King (1981b). The author was a member of the Joint Committee on Standards for Educational Evaluation for the last two years of the development of the revised standards (Joint Committee, 1994). However, the views expressed here are represented to be the views of only the author.

TM021610

BEST COPY AVAILABLE

ABSTRACT

Various strategies for optimizing the use of program evaluation information are recommended, based on a thorough examination of the literature. The recently revised *Program Evaluation Standards* (Joint Committee, 1994) are then reviewed as regards sensitivity to these recommendations. It appears that the new standards generally incorporate what the literature suggests should be regarded as "best practice".

A sense of frustration permeates the literature (King, Thompson & Pechman, 1981) on the use of program evaluation information. As Weiss (1972, p. 318) noted, evaluation's primary justification is that it contributes to the rationalization of decision-making. Although program evaluation can serve other functions, such as knowledge-building and theory-testing, unless it gains serious hearing when program decisions are made, it fails in its major purpose.

The purposes of the present paper are to review the literature on the use of program evaluations and to then correlate findings in the literature with emphases within the recently revised program evaluation standards (Joint Committee, 1981, 1994). The procedures employed to develop both editions of the standards, including those required for standards to be recognized by the American National Standards Institute as official ANSI-approved standards, are summarized by Sanders (1994). The revised standards themselves are described and summarized by Stufflebeam (1994).

I. Opposing Views of The Status of Use

The literature on the use of program evaluation information tends to be bifurcated. Views during the late 1970's and early 1980's tended to be pessimistic. Following this period, definitions of use tended to be broadened, and it was also increasingly recognized that not all program evaluations can or should be directly used in making specific decisions. At this point, the tenor of views of use then tended to become somewhat more optimistic.

Furthermore, views evolved that program evaluators themselves can and should take some responsibility for making program evaluations useful, and that such efforts can indeed be productive. These new views marked dramatic changes, because:

> Traditionally, the evaluator has been very hesitant to claim any responsibility for the use of his/her findings. This approach has helped make it very easy to ignore evaluation results. (Polivka & Steg, 1978, p. 697)

Concurrent with these shifts in perspectives, less empirical research regarding evaluation use was then reported in the late 1980's and the early 1990's.

I.A. Pessimistic Views About Evaluation Use

During the 1970's and early 1980's, many program evaluators came to be concerned that "there is something basically wrong with evaluation" (Orlandi & Conslave, 1977, p. 3). House (1973, p. 4) argued that "even under favorable circumstances evaluation data might account for only 20% of a decision." Guba and Lincoln (1981, p. ix) suggested that the failure to use evaluation findings has "almost assumed the proportions of a national scandal."

For example, Haenn (1980, p. 2) concluded that "evaluation results generally have served neither as a means of judging program results nor as a guide to program improvement." Williams and Evans (1969, p. 453) concluded that, "in the final analysis, the test of

the effectiveness of outcome data is its impact on implemented policy. By this standard, there is a dearth of successful evaluation studies."

Similarly, Alkin and Daillak (1979, p. 41) concluded that "there have been great hopes for evaluation, not only among evaluators themselves, but also among other educators, elected officials, and the public. Yet these hopes have dimmed." Worthen and Sanders (1973, p. 1) concluded that "evaluation is one of the most widely discussed but little used processes in today's educational systems."

Wholey, Scanlon, Duffy, Fukumoto, and Vogt (1970, p. 46) concluded that "the recent literature is unanimous in announcing the general failure of evaluation to affect decision-making in a significant way." Rippey (1973, p. 9) concluded that, "at the moment, there seems to be no evidence that evaluation, although the law of the land, contributes anything to educational practice other than headaches for the researcher, threats for the innovators, and depressing articles for journals devoted to evaluation." Finally, Stake (1973, p. 314) concluded that "we do not know whether or not evaluation is going to contribute more to the problems of education or more to the solutions."

This characterization apparently applied equally well to judicial (Saks, 1980) and legislative settings (Brandl, 1980; Mitchell, 1980). Unfortunately, several directors of evaluation from local education agencies (LEA's) during the 1970's also suggested that these characterizations may generalize to those settings too:

> In an ideal world we wouldn't have to worry about utilization. Educators would be eagerly awaiting our findings and would promptly rush to put them into practice. I don't need to tell you that isn't happening. (Holley, 1979, p. 2)

> The apparent nonuse of evaluation findings is one of the most vexing problems associated with the practice of evaluation in the public school setting. Even in districts with active evaluation sections it is difficult to establish a direct correspondence between evaluation results and educational decisions. (Novak, 1977, p. 1)

> All LEAs, with possibly a few exceptions, can point to their volumes of research and evaluation verbiage setting on the shelves of district administrators being used for little else than a door stop, swatting flies, or any of the other various and sundry purposes for which research is used in the public schools. (Kilbourne & DeGracie, 1979, p. 12)

Similar pessimistic views continue to be expressed more recently, though such views seem to be less typical of contemporary thinking. For example, Mitchell (1990) has suggested that,

> A worry for evaluators of public programs is the

> under-utilization of their work. One study after another has found that evaluations are sporadically used to improve policy outcomes, and in sundry instances, never even read. (p. 109)

Herman (1990, p. 1) recently suggested that, "In the face of such optimism about the power of data, we see repeated evidence that the actual impact of our work is quite modest." And Weiss (1987) has suggested that even evaluators who perform all the recommended practices to promote evaluation use have had indifferent success in making evaluation the basis of decisions.

I.B. More Optimistic Views About Evaluation Use

During this same period, some authors began suggesting that the extent of evaluation use may be underestimated in the literature (cf. Datta, 1978, p. 3). For example, Wise (1978, p. 24) argued that "if there is an evaluation utilization problem, it is not that decision-makers do not use the information they receive, it is that evaluators cannot easily see their information being used in the incrementalism of real-world decision-making."

Daillak, Alkin, and White (1978, p. 1) concurred, noting that "the few empirical studies that have been conducted, however, seem to present a picture which is less gloomy than many of the more speculative articles; they suggest that evaluation can have impact upon decision-making, although not necessarily the kind of dramatic go/no go influence some would wish."

There is empirical evidence that administrators do value evaluative information, although they might want changes in the methodology of some evaluations. For example, Alkin, Kosecoff, Fitz-Gibbon, and Seligman (1974, p. 28, emphasis in original) reported that "project directors found evaluations most useful in identifying possible problem areas... No project director indicated that he [sic] would prefer not to have an evaluator on the project."

More recently, Patton (1988, p. 10) emphatically argued, "I take strong exception to the assertion that evaluators who have followed conscientiously and skillfully the advise to work with intended users to achieve evaluation use have had 'indifferent success.'" Patton (1988) then cites specific examples of evaluation use.

However, one important empirical study does suggest that the results of use studies generally must be interpreted with some caution:

> The results of this study seemed at first to raise some doubt about the fundamental assumption underlying the study: the assumption that evaluation is an exercise in the service of decision-makers. While all those interviewed were, without exception, decision-makers in one capacity or another, their requests for reports were not always connected with the decisions they had to make. Frequently, they wanted the reports so as to inform the people above them or below them in the administrative hierarchy

> of their government agency. (Brickell, Aslanian & Spak, 1974, p. 56)

In any case, it is clear that the quality of utilization research and of estimates of use levels presumes an acceptable definition of use. As Tittle (1977, p. 3) noted, "papers concerned with the analysis of impact and impact assessment methodology have not always dealt with the problem of how to define impact." Patton (1978, p. 32) went so far as to argue that "the predominant image of nonutilization that characterizes much of the commentary on evaluation research can be attributed in substantial degree to a definition of utilization that is too narrow in its emphasis on seeing immediate, direct, and concrete impact on program decisions." Thus Alkin (1980a, p. 5) concluded that:

> From among these continuing strains of non-utilization, there now seems to be emerging a new methodology which points to instances in which evaluation information is in fact used. Moreover, these new results are not really contradictory with prior data on utilization; the [new] evidence on utilization rests upon a broader definition of utilization and different categories of evaluative information.

Similarly, Alkin, Daillak, and White (1979, p. 16, emphasis in original) argued that

> taken together, the studies and our observations and experiences suggest to us that evaluation <u>can</u> make a difference, that it does so more often than the published critiques suggest, that some school districts characteristically produce a high proportion of useful evaluations, and that some evaluators have acquired skills that allow them to carry out technically competent <u>and</u> programmatically influential evaluations.

In an important piece on utilization, Weiss (1979, p. 13, emphasis removed) argued that "until we resolve questions about the definition of use, we face a future of non-comparable studies of use and scant hope of cumulative understanding of how evaluation and decision-making intersect." Most of the agitation for an improved conceptualization of use stemmed from a growing recognition that use generally does not take very direct or dramatic forms.

As Alkin (1979, p. 3) suggested, "it is not enough to ask in September what the effects of the previous academic year's evaluation have been; as our illustration suggests, it may take two or three or more years before major program changes occur." Thus, Andrews (1979, p. 18) concluded that, "Of great importance is the finding that evaluations tend to have small incremental impacts on the programs; the 'big bang' theory of evaluation impact should apparently be discarded."

II. <u>Broadened Views of Use</u>

II.A. <u>Non-Use Can Be Tragic, In Some Circumstances</u>

From the beginning, program evaluators recognized that the non-use of evaluative information, when that use would be appropriate, can be tragic. For one thing, non-use represents an enormous waste of effort. As Datta (1979, p. 22) noted, "considerable effort is involved in conducting almost any evaluation: in identifying the evaluation question, in designing the study, in overcoming the obstacles to conducting an evaluation and protecting it methodologically from uninterpretability."

Non-use also represents the potential waste of substantial monies. For example, in 1974 direct expenditures on non-defense evaluation projects by the federal government alone amounted to $146 million (Kelezo, 1974). Surprisingly, "more and more money is being invested in evaluation studies at the same time that we are questioning their results and effects" (Raizen, 1978, p. 3). The indirect costs of non-use, when ineffective programs are not modified or discontinued, are even more staggering:

> The utilization of research crisis concerns the spending of billions of dollars in private and public funds to fight problems of poverty, disease, joblessness, mental anguish, crime, hunger and inequality. (Patton, 1978, p. 12)

However, the greatest tragedy of non-use, when use would be appropriate, is that the clients of educational and social programs receive less than optimally effective help. Failure to utilize evaluative information is tragic because, as Wise (1980, p. 16) noted, "no one else is given the resources and time to question, observe, assess, weigh, probe, and reflect that the evaluator is given."

II.B. The Rationality of Non-Use, in Some Circumstances

However, as program evaluators struggled with issues involving the use of program evaluation information, they increasingly acknowledged that the non-use of evaluative information can be quite rational, in some circumstances, for a number of reasons. That is, it has been increasingly recognized that the appropriate forms of use, or of non-use, are *circumstance-specific*.

For example, it became clear that <u>some evaluation studies are poorly done and do not merit use</u>. Ironically, there is empirical evidence (Alkin, Kosecoff, Fitz-Gibbon, & Seligman, 1974, p. 48) that poorer quality evaluations may be performed on projects which quality evaluations could most help to improve. At any rate, as Guba and Stufflebeam (1970, p. 6) observed:

> Many researchers make wrong assumptions about what an evaluation study should accomplish, and... [then] based on these erroneous assumptions, researchers foist bad advice upon unsuspecting and unsophisticated practitioners. As a consequence, evaluations are usually useless, and practitioners are largely justified in the jaundiced view they typically have taken about evaluation and its utility.

A common situation which justifies non-use occurs when

evaluators do not attend to a program's actual degree of implementation. As Guttentag and Struening (1975, p. 4) observe, "obvious though it may seem, evaluations continue without either raising or answering the primary question: 'Does the program [even] exist?'" Unfortunately, as Williams and Elmore (1976, p. xii) noted, "ignoring implementation has been equally disastrous for research and analysis." This situation continues, despite the availability of several strategies for measuring implementation prior to the comparison of results for program participants and non-participants (Revicki & Rubin, 1980), including most notably the model developed by Hall and Loucks (1977).

Administrators also often feel that evaluation studies do not merit use because the evaluation results contradict administrators' intuitions regarding program impacts. As Guba (1969, p. 1) noted, "for decades the evidence produced by the application of conventional evaluation procedures has contradicted the experiential evidence of the practitioner. Innovations have persisted in education not because of the supporting evidence of evaluation but despite it." This frequently occurs when summative evaluations find "no statistically significant differences" associated with a program--certainly a common result--but the practitioner's experience suggests that the program really did substantially alter classroom life. As Shapiro (1973, p. 527) argues, "while it is important to try to explain negative [i.e., statistically non-significant summative] test results, it is far more important to account for the disparity between the negative test findings and the clear differences observed in classroom behavior."

There is also evidence that non-use is rational, from the administrator's perspective, since the administrator may not view the world from within a scientific paradigm, and differences in the evaluator's and the administrator's perspectives may reduce trust and impede effective communication. As Deal and Rallis (1980, p. 216) explained:

> Theoretically, collaborative relationships require a shared perspective, high trust and power parity. The existing relationship between knowledge producer and user, however, is often characterized by different perspectives, low trust and an asymmetrical distribution of power.

Empirical research made clear how this situation might occur in public school settings:

> Not only have only 42% of them [LEA evaluation unit heads] not taught, but 70% have not run a school. This means that even when evaluation heads have teaching backgrounds, they do not take the typical advancement route to the central office. (Lyon, Doscher, McGranahan & Williams, 1978, p. 66)

There is also evidence that non-use may be rational when administrators invoke latent institutional or personal goals, which are quite distinct from formal program goals, as standards for determining program merit. For example, Granville (1977, p. 2)

explained that "a decision maker, in addition to considering whether or not a program has fulfilled its manifest objectives, must also consider its fulfillment of latent objectives, such as enhancing the agency's prestige or expanding its resources." It is also important to remember that administrators may have their own survival needs, and that these can affect their use of evaluative information:

> High level administrative tenure is quite short as is that of elected officials. Policy makers must demonstrably show actions in a short period of time as constituencies are not willing to wait. (Mathis, 1980, p. 2)

Finally, it is important, albeit painful, to acknowledge that some evaluation studies were never meant to be used. This is frequently the case when an externally mandated evaluation is involved (Burry, 1983). As Alkin (1976, p. 16) explained:

> Many practicing school administrators... believe that evaluation is simply an event that leads to compliance with various agency requirements. There is no real expectation that major basic decisions will be made. The name of the ballgame is simply not to get "dinged" by the governmental agency.

Alkin (1980a, p. 3) made the same point by way of analogy:

> And, to pursue the analogy, suppose the host at this garden party should insist that each of the guests periodically rate the quality of the party, or the drinks, or the food, etc.--it can't really be expected to have much impact. This somewhat peculiar, externally imposed requirement will be tolerated as part of the "price of admission," so to speak, but it won't really change the behavior of individuals.

Indeed, deliberate non-use is thoroughly rational if programs are not conceptualized well enough to possess "evaluability" (Rutman, 1977).

Taken together, these considerations suggest two important conclusions. First, we must not have unrealistic expectations for evaluation, at least as it is currently practiced (Daillak, Alkin, & White, 1978, p. 10). McLean (1979, p. 26, emphasis in original) helps to put this matter into perspective:

> Schools are overdetermined; that is, they are shaped by many forces, more even than are necessary to make them the way they are. Take away or change one force and nothing in a school may change.

Stevens and Tornatzky (1980, p. 340) concurred, although for a different reason. They argued that underutilization should be expected, since "underutilization of knowledge is actually quite common with an innovation such as program evaluation."

Second, and more importantly, it must be remembered that

> Evaluation has been ignored, misused, and overused, as well as appropriately used in policy and decision making. Our task is to make more appropriate uses of

evaluative information. (Braskamp & Brown, 1980, p. x)

Thus Caplan (1980, p. 5) argued that "there is a real danger in uncritically accepting utilization as desirable or in being oversold on its value. Not all utilization is good and not all nonutilization is bad." Similarly, Davis and Salasin (1975, p. 622, emphasis in original) argued that "it is a clear lesson that concern over utilization must emphasize appropriate use rather than just greater use."

II.C. Changed Views of Types of Use

The literature during the 1970s and 1980s reflected a growing recognition that subtle but still important types of use may be more typical than are direct uses of program evaluation information (Brown & Braskamp, 1980, p. 92). For example, as regards curriculum evaluation projects, Van den Berg and Hoeben (1984, p. 319) found that "at least a reasonable degree of use of evaluation results can be predicted." Duggan, Talmage and Rasher (1983) conducted interviews involving 26 evaluations in various settings, including school districts, and reported that 80% involved at least some client use of evaluation information. Thus, some LEA evaluators have come to express more optimistic views about evaluation use (Thompson, 1982).

King (1988, p. 287) summarized the situation by suggesting that "use is not unusual, but a rather frequent event, one occurring regularly throughout the evaluation process from its inception to well after its end." For example, Weiss (1977, p. 534) argued that, "government officials use research less to arrive at solutions than to orient themselves to problems... And [even] much of this use is not deliberate, direct, targeted, but a result of long term percolations of social science concepts, theories and findings into the climate of informed opinion."

As Leviton and Boruch (1983, p. 563) noted, "A truism of evaluation is that studies seldom contribute directly to identifiable decisions." But there are others forms of use.

Several conceptualizations of types of use have been offered. For example, Fullan (1979) suggested that information may result in changes in values, in understanding, in roles, in organization, or in materials. Some authors used different terms for types of use which apparently involve the same processes. As Weiner, Rubin, and Sachse (1977, p. 12) observed, "these categories are neither mutually exclusive nor exhaustive."

"*Instrumental*" use represents the more traditional view of use of evaluative information, e.g., the information results in "go/no go" decisions regarding program termination. It became recognized that this type of use rarely occurs, although some examples of "instrumental" use can certainly be identified (cf. Alkin, Daillak & White, 1979, p. 224). For example, Greene (1988a, p. 350) found that an approach emphasizing stakeholder participation in evaluation "can contribute to meaningful utilization of evaluation results, including, in the case studies reported herein, multiple instances of substantial instrumental use."

"*Conceptual*" use has been recognized as a much more common form of use. An example of conceptually targeted evaluation is provided by Cook (1974). Cook argued that "Sesame Street" should not be evaluated merely on the basis of the program's impacts on the reading readiness of children. It was suggested that the program's impacts on the gaps in reading readiness between lower socio-economic status children and other children should also be considered; this argument changed the conceptual frame of reference for evaluating that program.

Clearly, "conceptual" use of evaluation can have dramatic impacts at times. It is also clear that sometimes "it is difficult to determine where conceptual use ends and instrumental use begins" (Leviton & Hughes, 1979, p. 10).

"*Symbolic*" use has also been recognized as occurring commonly. Such use can take several forms. For example, program personnel who solicit evaluation only to satisfy external funding agencies are engaging in "symbolic" use; the evaluation is only used to persuade the agency that the game is being played according to the rules.

Lenihan (1977) provided an actual example of "symbolic" use. An evaluation demonstrated that several benefits could be realized by installing phones in a jail for inmate use. The evaluation's information was initially ignored.

> The [inmate] riot changed all that. When the time came, when overcrowding reached a breaking point, the knowledge produced from this research was put to use. It was not a sufficient cause for change, but in the end it did make a contribution (Lenihan, 1977, p. 583).

The evaluation provided a face-saving justification for then installing phones in the jail, after the riot, although the actual motives for doing so might have been less than scientific. Thus, it is perhaps not surprising that Knorr (1977) indicated that roughly 10% of surveyed administrators report they have used evaluations to legitimize decisions.

"*Ritualistic*" use is a misnomer, since this use can in fact have deliberate and important program impacts. Presumably evaluation has "anticipatory" affects on the behaviors of program personnel, because they know program processes and impacts are being measured. Surprisingly, the results of this type of use have not generally been empirically investigated.

Although these conceptualizations of use may accurately reflect "real world" evaluation dynamics, the shift from the more traditional view of use did produce some problems. An expanded view of use made it more difficult to study use phenomena. For example, "it is literally impossible 'to prove' [conceptual] use" (Fullan, 1980, p. 44).

Nevertheless, the trade-off of measurability in return for a more realistic perspective on use was probably worthwhile. As Braskamp and Brown (1980, p. viii) argued, "although the expanded definition makes utilization less dramatic and more difficult to explicitly measure and demonstrate, it represents a view of evaluation in which the role of human interaction in the

communication process is given more credence."

II.D. Changed Views of "Real World" Decision-Making

Just as the definition of evaluation use affected the assessment of the levels of use, perceptions of how decisions themselves are made also affected judgments of the quality of use. It was increasingly recognized that the nature of "real world" decision-making itself often precludes direct instrumental use of program evaluation information.

As Wise (1978, p. 6) explained, "referring to administrators as 'decision-makers' and to what they do as 'decision-making' may have been a first step in creating the utilization problem, for we expect to see decisions being made by someone called a decision-maker." Thus many evaluators initially presumed that evaluation ought to be used, because its evidence is rational.

However, as Weiner, Rubin, and Sachse (1977, p. 4) noted, "interestingly, their [evaluator's] recommendations are based upon a faith in rationality, not upon evidence concerning the factors influencing the utilization of evaluative information." As Caplan (1980, p. 4) suggested, "obviously, utilization is not a simple process--bureaucratic, ethical, attitudinal, and social considerations take precedence over the value of information in its own right."

Hayman (1979, p. 11) argued that the educational environment can be chaotic; consequently, administrators may only be able to invest careful and thorough thought in a few of their most critical decisions:

> [The "turbulent field" type of environment, one of four conceptualized environmental types,] appears to correspond to conditions facing education today. In a turbulent field, the accelerating rate and complexity of interactive effects exceeds the component systems' capacities for prediction and, hence, control of the compounding consequences of their actions.

Thus, Simon (1957, p. 204) suggested that administrators may be forced to engage in "*satisficing*", i.e., the process "of finding a course of action that is 'good enough.'" "Satisficing" is tenable because administrators believe they can change most courses of action if decisions later prove to be seriously mistaken.

Evaluators increasingly recognized that administrators at times may not behave "rationally", i.e., administrators' decisions may be rational only when they are viewed from within the administrators' frames of reference. Of course, it is also true that some administrators are less rational than others. For better or worse, some "decision-makers pride themselves on 'shooting from the hip' and would not have it any other way" (Guba, 1969, p. 17).

Administrators differ in their abilities to use information, and on personality factors that may impact predispositions to use evaluation information (cf. Brooks, 1984; King & Thompson, 1983a; McColskey, Altschuld & Lawton, 1985; Newman, Brown & Rivers, 1983, 1987; Newman, Bull, Brown & Rivers, 1986). Consequently, as

McClintock (1985, p. 19) found, "Administrators have highly varied perceptions of the usefulness of different information sources."

III. Empirical Studies on Evaluation Use

III.A. Overview of the Use Literature

During the 1970's and early 1980's a general consensus that there was a shortage of good empirical use studies emerged in the literature. For example, Davis and Salasin (1975, p. 626) reported that "a review of 1,200 references on [knowledge] utilization contained only 2 1/2% which pertained to evaluation, again even in the broadest sense." Furthermore,

> While much has been said and written about the problems besetting evaluation and about the underutilization of evaluation information, very few empirical studies of evaluation utilization have been conducted. Most of the literature is anecdotal in form. (Alkin & Daillak, 1979, p. 41)

Shapiro (1979, p. 1) agreed: "The literature on utilization, both applied and theoretical, tends to be ad hoc and nonrigorous." Cook (1978, p. 14) suggested that "the quality and imaginativeness of most (but not all) utilization studies leaves something to be desired."

Of course, it was difficult to conduct good utilization research. For example, as Stevenson (1979, p. 3) noted, "verbal acceptance of findings may not be followed by appropriate action. Verbal rejection of findings may be followed by actions which imply acceptance." Similarly, Barrios and Foster (1987, p. 9) observed that, "looking at implementation of evaluation recommendations alone may be misleading." Nevertheless, Caplan (1980, p. 9) may have overstated the situation when he said that "there are no tested propositions or even a substantial consensus as to what will work" in promoting use.

Fortunately, during the 1980's scholars worked to rectify these deficiencies. Noteworthy syntheses of studies in this genre were reported by Leviton and Hughes (1979, 1981), by Thompson and King (1981a, 1981b), by King and Thompson (1981, 1983b), by Cousins and Leithwood (1986), and by King (1988).

III.B. Factors Affecting Use

The literature does indicate that evaluation information is most likely to be used in either of two situations. First, evaluative information is most likely to be used when a program is novel and administrators cannot rely extensively on their experiential backgrounds in making judgments. For example, Matuszek and Holley (1977) reported that, in their experience,

> The [evaluation] office has tended to get the most response to its evaluation information when it really does represent something the decision maker doesn't already know. Thus, principals generally believe that they have a fairly good feel for teacher attitudes in their schools; a teacher attitude questionnaire is most likely either to

> confirm their feelings or to be rejected; student pattern of course choices (for example), on the other hand, is less likely to be at their fingertips.

An empirical study by Granville (1977, p. 6) corroborated this conclusion:

> The best way to express the magnitude of these effects is to say that the Political and Social Influence factors swayed principals by about one and a half intervals on a decision scale that ran across six intervals. The Objective Evidence factor had a separate effect of about one interval under the Novel program condition. Under the Routine program condition, as I mentioned, objective evidence had virtually no effect.

Second, the literature suggests that evaluative information is most likely to be used when only moderate changes in the program are required and the environment is not extremely politicized. As Meltsner (1976, p. 9) indicated, there are wide variations across programs regarding the environments in which programs operate:

> Sometimes the politics of a policy area are open and conflict ridden; sometimes they are closed, involving technical issues and technical men. Some policy problems have a dimension of crisis to them. Others are chronic; they never seem to go away and are subject to incremental solutions.

Consequently, Weiss (1972, p. 320) reported that "use of evaluation appears to be easiest when implementation implies only moderate alterations in procedure, staff deployment, or costs, or where few interests are threatened."

III.C. Strategies for Optimizing Use that May be Less Effective

The early literature included several recommendations for optimizing evaluation use which it now appears may not be particularly helpful. For example, some evaluators consider *formal evaluation models* (cf. Stufflebeam, Foley, Gephart, Guba, Hammond, Merriman & Provus, 1971) panaceas that can be used to create use. As Brown (1980, p. 4) noted,

> For a time, it was hardly respectable to be an evaluator without having your own model. You at least had to be a disciple of a proponent of a new model that was on the "cutting edge" in order to maintain some semblance of self-esteem. It is interesting to observe that there were very few wounds inflicted by that "cutting edge."

Due to these attachments to evaluation models, "the past few years have witnessed the development and proliferation of myriad evaluation models" (Thompson, 1980, p. 59). Indeed, this proliferation led to periodic efforts to synthesize the wide array of evaluation models (cf. Steele, 1973).

Today, however, it is clear that evaluation models such as the CIPP model (Stufflebeam et al., 1971), although important aids to

earlier efforts at conceptualizing evaluation, are less important when viewed as guides for evaluator behavior. Thus, based on his case studies of evaluative practice, Alkin (1979, p. 7, emphasis in original) found that "none of the five cases involved the application of a formal evaluation model." This finding was not an artifact of evaluator ignorance regarding the precepts of several evaluation models (Alkin, Daillak & White, 1979, p. 240).

More recently, it was also suggested (Dawson & D'Amico, 1985; Greene, 1988b) that evaluative information would be more widely used if evaluators, stakeholders, and administrators worked more closely together to *collaboratively formulate evaluation designs*. For example, Suchman (1972, p. 67) suggested that program goals must be explicitly specified prior to evaluation, because "unless a program can specify what value its activities are seeking to further, whether this be the amelioration of some specific social problem or the advancement of some broad humanistic goal, evaluation becomes meaningless."

Similarly, Patton (1978, p. 202) recommended collaborative development of evaluation designs, arguing that "it is crucial that identified decision makers and information users participate in the making of measurement and methods decisions so that they understand the strengths and weaknesses of the data--and so that they believe in the data." Finally, Ross (1980, p. 66) even suggested that evaluators and administrators should together specify decision rules in advance of program implementation; these rules were to specify what decisions would be taken if various evaluation results occurred.

Unfortunately, these strategies tend to look good on paper and work poorly in practice. For example, what Patton (1978, p. 100) termed the "goals shuffle" can readily destroy efforts to estimate discrepancies between program goals and actual program outcomes:

> The goals clarification shuffle involves a sudden change in goals and priorities after the evaluator is firmly committed to a certain set of measuring instruments and to a research design. The choreography for this technique is quite simple. The top priority program goal is moved two spaces to either the right or left and four spaces backward.

Rossi (1972, p. 229, emphasis in original) argued that what might be called the "methodology shuffle" can also occur if evaluation results prove to be unpopular:

> It is easy to attack the methodology of any study: methodological unsophisticates suddenly become experts in sampling, questionnaire construction, experimental design, and statistical analysis, or borrow experts for the occasion.

Of course, decision rules tend to become obsolete once either goals or methodology shuffles have been performed.

In summary, the collaborative identification of goals, methods, and even decision rules is all well and good. However, it is important to recognize that these strategies are most likely to work in situations in which evaluation use is least threatened.

More importantly, it must be recognized that these strategies do not themselves directly address the primary factors that apparently affect use. Rather, these strategies are indirectly instrumental in promoting the factors that do appear to directly promote use.

III.D. Strategies for Optimizing Use that May be Effective

The literature also included some recommendations for optimizing use which, although they are plausible and noteworthy, are easier said than done. For example, Havelock (1968) suggested that use will be optimized if someone performs a "*linking agent*" function. Hayman (1979, p. 1, emphasis removed) defined "linkage" as "a process of promoting knowledge utilization in educational organizations, and a 'linking agent' is an individual or group which causes linkage to occur."

It may be that evaluation is most likely to occur when an administrator who is unassociated with the evaluation, i.e., is perceived by fellow administrators as being objective, "adopts" an evaluation study and begins pushing for implementation of the study's results. The dilemma, of course, is that to be most credible this process must be spontaneous, so there are no guarantees that this form of linkage will occur. However, this form of linkage may be more likely to occur if the evaluator targets results toward a larger number of administrators.

It has also been suggested in the literature that utilization will be optimized if evaluative information is presented to administrators in a *timely* fashion. Randall (1969, p. 1) portrayed what may be a common situation:

> There is a timeworn and oft-recurring spectacle of the frantic but finally productive researcher-evaluator, who rushed into the executive offices with his [sic] data analysis finally complete, his report prepared and in hand, only to find that the executives, several months previously, had made the important decisions that locked up the monies and committed the organization for the ensuing months ahead.

Unfortunately, it is not always possible to anticipate when information will be needed in service of decision making. In fact, as Brickell, Aslanian, and Spak (1974, p. 24) noted, "he [the administrator] can never know when he will need it [evaluative information]. The process of government decision-making is not so orderly or regular that he [sic] can schedule his need for information." It is also important to recognize that timeliness is important for "instrumental" use, but may not be so for other types of use (Young & Comptois, 1979).

Finally, Johnston (1978, p. 1) has suggested that it is important to *target evaluation toward identified administrators*.

> There is a sort of ecology for each educational program, a network of people in different roles who influence (or are influenced by) the outcome of the program being evaluated. If this is true, and research utilization is the goal of the evaluator,

> then there are multiple audiences for an evaluation, not just the decision-maker who commissioned the evaluation. So the evaluator has a first task of identifying who these other actors are.

However, this effort can be frustrated by the complexity of the organizational network.

As Randall (1969, p. 7) explained, "typically, the decision process in an organization involves a complex network of persons who have varying degrees of influence on the one who may have constituted authority to make any given decision." The situation is further complicated because, as Granville (1978, p. 29, emphasis in original) noted, an evaluation study "has to persuade not just the people who ostensibly make the decisions, but also the people they have to persuade." Thus, Alkin and Kosecoff (1973, p. 3) concluded that "identification of the program's decision maker(s) is perhaps the most elusive variable associated with a decision context."

III.E. Strategies for Optimizing Use that Appear Essential

The literature on evaluation use includes several recommendations for optimizing use that appear to generally be essential to optimizing evaluation use. These recommendations involve (a) identifying evaluation issues, (b) acknowledging evaluation subjectivity, (c) considering political realities, (d) explicitly recommending policy decisions, (e) not overemphasizing single forms of proof, and especially (f) building personal rapport with administrators and program personnel.

Recommendations for seven specific practices are embedded within this discussion. However, in considering these seven recommended practices, it is important to remember that the factors affecting evaluation use emerge as "...discernible patterns, but they are highly complex and highly interactive.... Each evaluation context must be viewed as unique" (Brown, Newman & Rivers, 1985, p. 444).

Furthermore, it is crucial that the evaluator use a *holistic approach* to adopting these strategies. As Weiner, Rubin, and Sachse (1977, p. 23) argued, "attempts to increase evaluative influence which focus on a few of these factors in isolation and which do not recognize the highly complex and interactive system of forces constraining evaluator activity are likely to fail to alter the overall effects of the system." Patton (1978, pp. 19-20) concurred, noting that "the overall problem of underutilization of evaluation research will not be solved by compiling and following some long list of evaluation proverbs and axioms."

III.E.1. Issue Identification

Evaluation results will enjoy "instrumental" and "conceptual" use only if the results address *issues* of concern to administrators. As Alkin, Daillak, and White (1979, p. 238) noted, "if the evaluation addresses a pressing concern of a potential user, then the evaluation information is more likely to draw, and hold, the user's attention." Thus Patton (1978, p. 83) went so far as to suggest enhancing "utilization by focusing on fulfilling one

purpose extremely well, so that at least the decision makers' central questions are answered." Similarly, Alkin and Daillak (1979, p. 47) argued that "evaluators who concentrate on the mandated evaluation tasks run the very real risks of losing the local audiences."

However, it may be difficult to identify the issues that are most important to administrators. For example, in a unique project designed to be particularly responsive to administrators' information needs, Fletcher (1972, p. 15) found that "like teachers, administrators could not often identify kinds of data they could use. And in many cases the kinds they wanted were totally beyond our capacity to provide." Furthermore, the situation which Ingison (1979, p. 2) observed at the federal level may also occur in local education agencies: "At the [National Science] Foundation (and elsewhere at the federal level, I suspect), the pressure is always on to get the evaluation study set up and brought in quickly (preferably yesterday)." Despite these difficulties, to maximize the likelihood that evaluative information will be used, it is recommended that

1. *Evaluators should concentrate evaluation efforts on the highest priority information needs of specific administrators, even if these needs require work beyond that mandated by external funding agencies.*

Since administrators are not always able to anticipate or articulate future information needs, evaluators "should anticipate questions and be proactive" (Law, 1980, p. 74). Stake (1973, p. 305) made a similar point: "The evaluator, I think, has a responsibility to snoop around and to guess at what decisions may be forthcoming. He [sic] should use these guesses to orient his evaluation plan." Gorham (1970, p. 104) argued that evaluators should "be clairvoyant about forthcoming issues." Meltsner (1976, p. 127) cited the example of "one analyst [who] likes to follow his client around for a week and attend the meetings he does, and the like, to see what 'is hurting the client.'" These arguments suggest that

2. *Evaluators should identify some evaluation issues on the basis of emphatic and proactive anticipation of administrators' future information needs.*

If these anticipated needs do not arise, evaluators need not highlight the results of the inquiries that they initiated. In any case, evaluator credibility should be improved when administrators sense a sincere effort to be responsive, even though this anticipation will not always be precisely accurate.

In addressing evaluation issues, evaluators must also take into account the factors that determine the perceived salience of information. For example, as Brown (1973, p. 2) explained, "the question as to when and which levels of information school management needs depends on the length of time a program has been in operation and the degree to which that project overlaps other programs within the school system." An administrator's location within the organizational hierarchy also affects extent of felt need for information (Braskamp, Brown & Newman, 1978, p. 449).

Furthermore, organizational position affects the type of information that is required; "those officials who are in a position to control the project from day to day or month to month and who are responsible for exercising such control have a far greater interest in monitoring reports than decision-makers at higher executive levels" (Brickell, Aslanian & Spak, 1974, p. 59).

Evaluators should also remember that "research is often most useful to those who do not have the authority to promote a policy, i.e., teachers" (Hamilton, 1980, p. 7). In short, evaluators must tailor information studies to meet the different needs of various evaluation audiences. This may entail steps such as writing several versions of reports or conducting "extra" evaluation inquiries.

III.E.2. Acknowledging Subjectivity

At least traditionally, many evaluators liked to believe that evaluation is objective, apolitical, and appropriately empirical. All three of these beliefs can take the form of myths that hinder evaluation use. For example, Wilensky and Lebeaux (1958, p. 20) observed that "what the social scientist thinks of as 'objective investigation' the practitioner often takes as 'hostile attack.'" As Goodrich (1978, p. 632, emphasis in original) suggested, "what has happened is that we have tried to avoid the phenomenon of subjectivity in order to avoid the charge of subjectivity."

However, Patton (1978, p. 237) argued that:

> The fundamental issue is whose values will bias the question, not whether or not questions will be biased. In a very real sense all questions are biased, but biased questions can be either open or loaded.

Meltsner (1976, p. 261) put the matter of *acknowledging subjectivity* nicely into perspective:

> Trust is also nourished by the analyst's attempting to be objective. This does not mean that either the analyst or the analysis can be objective in an absolute sense. Both do have their values. What it does mean is that the analyst should attempt to give as straight an answer as he [sic] can.

Some administrators perceive evaluation as a two-edged sword. The pretense of objectivity imbues evaluation with credibility. The administrator who is confronted with an "unfavorable" result knows that any result can be attacked on several grounds. The administrator who is confronted with a "favorable" result may perceive the result as a weapon against adversaries. However, firm stances may not be taken until the results dictate them. To reduce some administrators' rather schizoid views of evaluation, and for the mutual benefit of all concerned,

3. *Evaluators should acknowledge the subjective elements in their work, when these elements are unavoidable, and offer their work merely as informed but not omnipotent support for decision-making.*

However, this recommendation must be implemented with care, because sometimes administrators misinterpret honest caveats as admissions

of avoidable and serious evaluation flaws.

III.E.3. Considering Political Realities

Although many evaluators are still not comfortable admitting it, evaluation cannot avoid being political (Weiner, Rubin & Sachse, 1977, p. 19). Evaluators must recognize *political realities* and consciously work within the context of these realities.

Admittedly, some evaluators find such an orientation disconcerting and somewhat foreign. As Patton (1978, p 46) explained:

> The traditional academic values of many social scientists lead them to want to be nonpolitical in their research. Yet they always want to affect government decisions. The evidence is that they cannot have it both ways.

Isaac (1980, p. 3) concurred:

> Programs that are politically conceived and implemented, are also sustained and defended politically. It seems the better part of wisdom neither to be surprised nor offended by this phenomenon.

According to Cohen (1972, p. 139), "to the extent that information is an instrument, basis, or excuse for changing power relationships within or among institutions, evaluation is a political activity."

The implications of this situation have been identified by Dickey (1979, p. 3):

> Judgment has an awesome ring, and it is not surprising that those who are being judged feel anxious, even threatened. Add to this the political context in which the process takes place (and there is always a political context) and we have all the ingredients for dysfunctional communication--high levels of stress leading to communication patterns arising from individual defense mechanisms.

This does not mean that evaluators must themselves participate in political activity. However, as Meltsner (1976, p. 43) suggested, the effective evaluator "tries to understand political considerations and then to make them an integrated and explicit part of his [sic] analysis." According to Brown and Braskamp (1980, p. 93), "this means that the relationship between the evaluator and key program staff, and the evaluator's understanding of the organization in its internal and external political environment, are critical for successful utilization." This suggests that

4. *Evaluators should understand the politics of their agencies and attempt to meet the political needs of involved persons whenever doing so will not jeopardize the integrity of the evaluation.*

III.E.4. Explicitly Recommending Policy Decisions

If evaluation is not a purely objective and also is inherently political, then it is reasonable to expect evaluators to offer or at least consider policy alternatives within their works. Haller

(1974, p. 403) observed that "evaluation problems concern decisions. Decisions presume the existence of alternatives, and so the purpose of evaluation is to help delineate alternatives and to provide information to help decision makers arrive at more rational choices."

However, Weiss (1979, p. 3) noted that many evaluators are not sufficiently policy oriented; "evaluators do not always--or even often--come up with data that give explicit guidance for action." As Zepeda (1980, p. 1) noted, this situation is not entirely satisfactory: "Informing local decision makers that the local Title-I program is not effective does not give them the information that they need to improve it." The consequences of this situation were explained by Deal and Rallis (1980, p. 210): "By only describing in a controlled form what already exists, scientific research does not directly promote learning for the craftsman. It might even help to maintain the status quo."

Some empirical research has indicated that some evaluators do delineate policy choices and make policy recommendations. For example, Alkin, Kosecoff, Fitz-Gibbon, and Seligman (1974, p. 19) found that "thirty-nine of the 42 evaluators made recommendations for project modification." But this result contradicts other research (Goldberg, 1978, p. 16) which found that administrators do not perceive evaluation to be too helpful in "finding possible courses of action and choosing among alternative courses of action." This analysis suggests that, in evaluation reports, "reasonable alternatives must be provided. Such alternatives have a good chance of being tried when provided" (Ingison, 1979, p. 4). In short, evaluation might be used more effective when program evaluators explicitly delineate policy alternatives and even make policy recommendations.

Isaac Assimov (1972) wrote a short story that illustrates the psychology of this situation. A young scientist proves mathematically that an energy policy is going to result in the destruction of the universe. However, no one will believe his proof until decades later a mechanism for still generating huge quantities of energy without destroying the universe is also presented. The moral is that administrators may be more likely to act on evaluation information when they are presented with some policy alternatives to consider.

Will administrators perceive recommendations as an unwarranted intrusion into the policy arena? They will if the evaluator's offerings take the form of grandiose schemes. They may not if specific policy alternatives are mentioned and the evaluator merely presents objective evidence, both pro and con, regarding the best predictions about likely program impacts.

Some evaluators perceive administrators as being very jealous of their decision-making turf. In some cases these perceptions are fully justified. But many administrators do <u>not</u> feel threatened by decision alternatives and policy recommendations. Choosing is the business of decision-making, so most administrators do not fear either choices or recommendations, especially if they believe that recommendations are being offered in a sincere attempt to help as

against being offered in a sincere attempt to be Machiavellian. Thus,

5. *Evaluators should delineate policy alternatives and, where appropriate, also recommend particular policy decisions.*

It may even be reasonable in some situations to view the conveying of evaluation information as persuasion (Eason & Thompson, 1988; Thompson, 1981). However, as Newman and Brown's (1980) results indicate, the utility of these efforts will partly be determined by the situation-specific personalities and needs of the involved administrators.

III.E.5. Not Overemphasizing Single Forms of Proof

Some administrators have a stereotypic view of evaluators in which evaluators are perceived as "technicians." Meltsner (1976, p. 23) summarized the stereotype thusly:

> Unlike the messenger of ancient times, the technician does not fear for his head when he [sic] has to bring bad news. No, he would not soften his findings; he would do "the best analytical job that can be done in conformity with the principles of economics [or his discipline]." He refers to his work as "honest analysis," and he complains about analysts having to provide justification for a decision that has already been made.

However, the evaluation features that are persuasive to administrators stand in stark contrast to what the technician will emphasize. As Leviton and Hughes (1979, p. 23) suggested, "administrators prefer qualitative information to the quantitative data that evaluators frequently supply."

For example, Oman and Chitwood (1984, p. 303) conducted two hour interviews with 50 evaluators and administrators, and reported that, "Studies that used structured and unstructured interviews [for data collection] appeared to have a higher-than-average level of acceptance." Similarly, Turner, Hartman, Nielsen and Lombana (1988, p. 129) found that "a variety of data gathering methods not only affects uniquely the use of evaluation information but contributes to related utilization factors." Thus, it is clear the evaluators should generally not present *single forms of proof* involving only quantitative information.

Empirical research supports the view that administrators often prefer qualitative information over quantitative information. Alkin's (1980b, p. 24) well-known naturalistic utilization studies yielded the conclusion that "little evidence was found in the case studies that research rigor was an important factor affecting utilization." Simulation research by Brown and Newman (1982, p. 201) was even more dramatic regarding this point:

> The simple addition of an inferential statement, such as "these differences were statistically significant at the .05 level" however, resulted in lower levels of agreement [with policy recommendations]. In fact, for three of four recommendations, the inclusion of the inferential

statement resulted in levels of agreement lower than in the No Data [experimental] condition.

However, it is important to note that the use of data does interact with other result features in determining audience reaction (Brown, Newman & Rivers, 1980, p. 72), so a simple interpretation of these results is not possible. Of course, some administrators' disdain for quantitative empiricism is rational if we acknowledge that quantitative forms of representation "inherently are insensitive to some of the significant aspects of classroom life" (Eisner, 1980, p. 11).

This is not to suggest that evaluators should forego the quantitative aspects of their work; rather "the central message in this regard is that it is not enough to conduct methodologically sound research" (Johnson, 1978, p. 12):

6. *Evaluators should generally report both quantitative and qualitative evaluation data in their work.*

The two foci reinforce each other. The emphasis on qualitative data communicates both an understanding of program ecology and a commitment to program improvement; the previous discussion has suggested that these messages are essential. An emphasis on formative process data should then help make quantitative summative results more credible. Of course, quantitative data must still be presented in concrete and understandable terms. In any case, evaluators might also do well to be sure that they understand the meaning and the limits of their inferential methods (Carver, 1993; Thompson, 1989, 1993).

III.E.6. Rapport with Stakeholders and Administrators

The preceding prescriptions speak to the personality of the evaluator. Such prescriptions are unavoidable. The literature suggests that the most critical determinant of evaluation utilization is what Patton (1978) has termed "*the personal factor.*" As Cronbach et al. (1980, p. 6) argued, "nothing makes a larger difference in the use of evaluations than the personal factor--the interest of officials in learning from the evaluation and the desire of the evaluator to get attention for what he knows."

This emphasis has been stronger in more recent research investigating evaluation use. For example, Pflum and Brown (1984, p. 42) reported that their results "...tend to support the assertion that use of evaluation information is influenced by 'personal factors.'" Similarly, Brown, Newman and Rivers (1985, p. 444) reported that, "The findings are also congruent with the assertion that evaluation use is influenced by 'personal factors'."

However, to some degree these influences may be situation-specific. For example, Newman, Brown, Rivers and Glock (1983, p. 118) studied school board members, and found that "Subjects were more likely to want to make informal contacts when making decisions about high conflict situations."

Nevertheless, Oman and Chitwood (1984, p. 304) conducted interviews with 50 evaluators and administrators, and reported that "The level [of use] was much lower when these [evaluation] processes were handled by teams of analysts rather than by

individual analysts." Similarly, Ripley (1985, p. 424) conducted a simulation study involving written reports, as well as both video and audio cassette versions of evaluation reports, and found that, "Persons who receive information through a nonwritten source are more likely to agree with the evaluator's recommendation."

Evaluators sensitive to the personal factor will respond in two ways. First, following the suggestion of Patton (1978), evaluators will identify the evaluation's relevant administrators and stakeholders, and second, they will continually work on affecting utilization throughout the course of the evaluation.

Several writers have noted the importance of good evaluator-client relations:

> The evidence on dissemination suggests that informal communication that cuts the red tape may enhance utilization, although quality of information may sometimes suffer and dissemination will be haphazard. (Leviton & Hughes, 1979, p. 21)
>
> For while information is an essential resource for decision makers, the manner in which it is converted into policy is based as much or more on interpersonal, organizational, and psychological factors than on the actual information itself. (Guskin, 1980, p. 45)
>
> Utilization is usually the result of the relationship between the evaluator and the user more than anything else. If the user knows and respects the evaluator, utilization has its highest potential. (Holley, 1979, p. 8)
>
> The major barriers to successful evaluation are not technical and methodological, though these are certainly important and worthy of further effort, but are rather the structural constraints and requirements and the interpersonal relationships which characterize the evaluation endeavor. (Gurel, 1975, pp. 27-28)

To be effective,

7. *Evaluators must demonstrate to stakeholders and administrators that they sincerely care about the needs of program staff and the program's clients.*

An example may illustrate how this admonition can work in practice. Several years ago the present author was working in a local education agency as an evaluator of a court-ordered magnet school program. Since one of the program's functions was to promote racial integration of students, each magnet school was charged with actively recruiting students representing various ethnic backgrounds.

After several months of operation it became clear that the staff of one project had deliberately avoided this responsibility in order to gain time to get their program on the ground. Although

this decision was perfectly rational from the perspective of the program's staff, in light of the court-order this inactivity could have had serious implications for the school district. If "breathing time" was required, then the admissions phase of the program could be delayed, but it was vital that recruiting activity begin as soon as possible so that lag times would be minimized.

Consequently, it was necessary to meet with the program director and the administrator who supervised the director. The evaluator took this position:

> This is a potentially serious matter which I must present in my next process evaluation report. But how can I help you? Since your program has academic admissions criteria, would you like me to analyze the district's computer tapes from the last system-wide testing and generate mailing labels for eligible students in the grade levels you serve? You could write them and invite them and their parents over to see what your program has to offer. This way when I discuss your recruiting activities I can point to concrete evidence that the situation is being corrected.

It was not technically the evaluator's job to generate mailing labels. However, actions like these are important because they communicate the evaluator's concern for the program and its stakeholders.

The "personal factor" in part means that evaluators try <u>not</u> to be threatening and authoritative; how evaluators comport themselves affects the psychological frameworks with which administrators interpret evaluative information. This conclusion may discomfort some evaluators who believe that they offer objective truth which they believe should have intrinsic value independent of evaluator personality or approach. Nevertheless, administrators have their own paradigms for viewing the world, and these paradigms are rational to them. Evaluators must accept that the manner in which the evaluator interacts with administrators and stakeholders will affect the credibility that the administrators and stakeholders vest in subsequent evaluation results.

IV. <u>Correlation of the 1994 Standards with the Use Literature</u>

IV.A. <u>Relevance of the Use Literature to the Standards</u>

The Joint Committee (1994, p. xviii) has suggested that, "Taken as a set, the 30 [program evaluation] standards provide a working philosophy for evaluation." Throughout their book the members of the Joint Committee repeatedly focus on <u>use</u> as the criterion against which to judge the value of program evaluations. Thus, the empirical literature on evaluation use is relevant to commentary on the validity of the Joint Committee's various recommendations.

A representative cascade of quotations from throughout the 1994 volume indicates the Joint Committee's concern with the use of program evaluation information:

> Education and training programs are evaluated in

order to determine their quality and gain direction for improving them. (Joint Committee, 1994, p. 1)

[The] utility standards guide evaluations so that they will be informative, timely, and influential. (Joint Committee, 1994, p. 5, emphasis added)

...the evaluator will want to know what was done in the evaluation and what came out of it, as well as the evaluation's impact. (Joint Committee, 1994, p. 11, emphasis added)

...the Committee is confident that the standards will lead to sound, useful, ethical, and cost-effective evaluation... (Joint Committee, 1994, p. 20, emphasis added)

Prospects for acceptance and use of the evaluation would have been enhanced if the evaluators had taken steps to ensure their credibility... (Joint Committee, 1994, p. 33, emphasis added)

Evaluators should communicate evaluation findings to intended users at times when the information can best be used. (Joint Committee, 1994, p. 53, emphasis added)

[standard U7] Evaluations should be planned, conducted, and reported in ways that encourage follow-through by stakeholders, so that the likelihood that the evaluation will be used is increased. (Joint Committee, 1994, p. 59, emphasis added)

Evaluations should be expected to withstand the critical examination of those whose lives they may affect and to provide them with useful information. (Joint Committee, 1994, p. 109, emphasis added)

The summative role of metaevaluation is to assess the worth and merit of a completed program evaluation. In this role, metaevaluation addresses such questions as: ...Were they [the program evaluation's findings] used? (Joint Committee, 1994, pp. 185-186, emphasis added)

Indeed, it is even suggested that program evaluations for which use cannot be predicted should be terminated:

Metaevaluation can serve the valuable function of helping prevent or terminate while still in progress a program evaluation whose results predictably would never be used. (Joint Committee, 1994, p. 186, emphasis added)

Use is also consistently one justification offered for attention even to proprietary standards. For example, the Joint Committee (1994, p. 93) noted that:

> In addition, [program evaluation] conclusions and recommendations may be discounted if it is learned they were derived from information obtained illegally or unethically.

With respect to misappropriation of evaluation funding, the Joint Committee (1994, p. 121) argued that, "In addition, alleged misuse of funds can be used in attempts to discredit an evaluation." And the Joint Committee (1994, p. 177) noted that, "If stakeholders do not receive sufficient information for determining whether the [program evaluation] conclusions are warranted, they may disregard them."

Throughout much of the Joint Committee's discussion, attention often seems to be focused on the kinds of direct, *instrumental* use that the literature suggests are not as frequent as other types of use. For example, the Joint Committee (1994, p. 6) noted that,

> These [accuracy] standards are intended to ensure that an evaluation will reveal and convey accurate information about the program's merit and/or worth.

And subsequently the Joint Committee (1994, p. 185, emphasis added) noted,

> Program evaluation is difficult to do well but may provide critical support to effective programs or result in <u>modification or cancellation</u> of ineffective programs.

However, at one point the Joint Committee does acknowledge that other forms of use are both expected and legitimate:

> The impact of an evaluation refers to the influence it has on the decisions and follow-up actions of members of the audience. It also refers to the <u>conceptual influence</u> it has on stakeholders. (Joint Committee, 1994, p. 59, emphasis added)

In the aggregate, it appears that the Joint Committee's focus was on use, including instrumental use, but was not necessarily limited to this one form of use. However, the Committee's premise seemed to be that if instrumental use is not at least considered by program evaluators, then such use is less likely to ever occur.

IV.B. <u>Some Recommendations Correlated to the Use Literature</u>

Previously, six classes of strategies for optimizing use that appear essential were discussed. The 1994 standards can be evaluated within each of these same six categories.

IV.B.1. <u>Issue Identification</u>

It has been suggested that proactive issue identification is critical to responsive evaluation. One emphasis in this treatment was on the anticipation of administrators' and stakeholders needs, even in cases where these individuals may not recognize or anticipate certain needs.

The standards do clearly consider the important concerns of

issue identification. For example, the Joint Committee (1994, p. 25) noted that:

> If stakeholder identification [standard U1] is not done, the evaluation may become a misguided, academic exercise, the results of which are ignored, criticized, or restricted because they do not address anyone's particular questions.

However, the treatment generally presumes that administrators and stakeholders always know what information they will need, if evaluators only ask them to ponder these potential needs. This treatment does not patronize. A presumption is made that administrators and stakeholders generally are in the best position to judge their own needs and interests. However, it should also be noted that in some instances prudent program evaluators are able to anticipate what would otherwise be unforeseen information needs.

IV.B.2. Acknowledging Subjectivity

It has been suggested that program evaluation is inherently, to at least some degree, not a completely objective business. The standards unquestionably recognize this reality. For example, at one point the Joint Committee (1994, p. 37) noted that, "This weeding-out procedure [involving de-emphasizing less important information] requires judgment."

Elsewhere, the Joint Committee (1994, p. 38) argued that, "Evaluators, like other professionals, bring their own preferences to the task of carrying out an evaluation." And the Joint Committee (1994, p. 44) recognized as a "common error", "A. Assuming that evaluations can be objective in the sense of being devoid of value judgment."

IV.B.3. Considering Political Realities

It has been suggested that program evaluators must recognize that program evaluations are conducted in a political context. Again, the Joint Committee clearly accepted this view. The Joint Committee (1994, p. 4) noted,

> The standards also help evaluators identify and confront political reality. Political agendas and money are sources of power that may corrupt evaluation in any setting.

Indeed, the Joint Committee (1994, p. 71) even suggested that, "...evaluators who are sensitive to political pressures sometimes will be able to make constructive use of diverse political forces in achieving the purposes of the evaluation."

IV.B.4. Explicitly Recommending Policy Decisions

It has been suggested that program evaluators generally ought to delineate policy options and even advocate particular policy choices. The Joint Committee clearly recognized that program evaluation serves decision-making. However, the tenor of this discussion was generally framed as providing information, rather than as recommending particular choices. Of course, the appropriateness of recommending specific policy choices varies

considerably from setting to setting and from evaluation to evaluation.

IV.B.5. Not Overemphasizing Single Forms of Proof

It has been suggested that program evaluators should not overemphasize single forms of proof. This view is embedded throughout the Joint Committee's (1994) book. For example, the Joint Committee (1994, p. 4) noted that, "In addition, the standards encourage the use of a variety of evaluation methods." Elsewhere the Joint Committee (1994, p. 141) argued,

> It is desirable that information be obtained from a variety of sources, so that the information from different sources can be compared for congruity or added perspectives.

IV.B.6. Rapport with Stakeholders and Administrators

Particular emphasis was placed here on the "personal factor", i.e., building rapport with stakeholders and administrators. Similar views are emphasized throughout the standards volume.

For example, standard U2 requires that,

> The persons conducting the evaluation should be both trustworthy and competent to perform the evaluation, so that evaluation findings achieve maximum credibility and acceptance. (Joint Committee, 1994, p. 31)

The Joint Committee (1994, p. 31) also suggested that:

> When conducting an evaluation, evaluators should maintain a pattern of consistent, open, and continuing communication and approachability with their clients and other stakeholders while still offering expertise and maintaining impartiality.

V. Summary

Various strategies for optimizing the use of program evaluation information have been recommended, based on an examination of the literature. The recently revised *Program Evaluation Standards* (Joint Committee, 1994) were reviewed as regards sensitivity to these recommendations. It appears that the new standards generally incorporate what the literature suggests should be regarded as "best practice". Hansen (in press) and Patton (in press) seem to concur in this view.

Patton (1978, p. 96) has noted that "increasing utilization potential does not guarantee utilization of findings. There are no guarantees." Nevertheless, as King (1988, p. 287) has observed:

> Of the four groups described here, evaluators probably have the greatest potential for improving the use process because, unlike the others, they already have a thorough knowledge of the evaluation process.

Thus, it is probably not surprising that the revised standards place the onus of considerable responsibility for promoting use on the shoulders of evaluators themselves. The Joint Committee (1994,

p. 59) argued that, "Evaluators must not assume that improvements will occur automatically once the evaluation report is completed."

It is encouraging to note that concerted efforts in these directions can indeed yield desired effects. For example, in one study Huberman (1990, p. 386) recently reported that, "...Dissemination effort (aggregated) correlates at .70 (rho) with conceptual use and at .46 (rho) with instrumental use."

Burry (1985, p. 14) nicely characterizes the behaviors and attitudes that evaluators can employ to optimize the use of program evaluation information:

> The evaluator who adopts the use-promoting stance suggested above takes an important step toward fostering the trust and harmony that underlie rapport with users, a rapport that is further strengthened when the evaluator is sensitive to the program's political dynamics and understands that evaluation information is only one of many possible inputs to the decisionmaking process and that people with different attitudes, backgrounds, and power or prestige are likely to contribute to that process.

The revised standards (Joint Committee, 1994) seem to incorporate similar views.

References

Alkin, M.C. (1976, October). The title of this speech is... Paper presented at the annual meeting of the California Educational Research Association, Burlingame, CA. (ERIC Document Reproduction Service No. ED 160 652)

Alkin, M.C. (1979, April). Using naturalistic research for the study of evaluation utilization. Paper presented at the annual meeting of the American Educational Research Association, San Francisco. (ERIC Document Reproduction Service No. ED 170 378)

Alkin, M.C. (1980a, April). Improving Title I technical assistance: An evaluation utilization perspective. Paper presented at the annual meeting of the American Educational Research Association, Boston.

Alkin, M.C. (1980b). Naturalistic study of evaluation utilization. In L.A. Braskamp & R.D. Brown (Eds.), Utilization of evaluative information. San Francisco: Jossey-Bass.

Alkin, M.C., & Daillak, R.H. (1979). A study of evaluation utilization. Educational Evaluation and Policy Analysis, 1, 41-49.

Alkin, M.C., Daillak, R.H., & White, P. (1979). Using evaluations: Does evaluation make a difference? Beverly Hills: Sage Publications.

Alkin, M.C., & Kosecoff, J.B. (1973, April). A new eclectic model for the redirection of evaluation efforts. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Alkin, M.C., Kosecoff, J.B., Fitz-Gibbon, C., & Seligman, R. (1974). Evaluation and decision-making: The Title VII experience. In CSE Monograph Series in Evaluation (No. 4). Los Angeles: Center for the Study of Evaluation.

Andrews, J.V. (1979). Reactions to program evaluation: A qualitative analysis (Doctoral dissertation, Cornell University, 1979). Dissertation Abstracts International, 39, 6717A. (University Microfilms No. 79-10,731).

Assimov, I. The gods themselves. New York: Doubleday, 1972.

Barrios, N.B., & Foster, G.R. (1987, October). Utilization of evaluation information: A case study approach investigating factors related to evaluation utilization in a large state agency. Paper presented at the Annual Meeting of the American Evaluation Association, Boston, MA. (ERIC Document Reproduction Service No. ED 292 814)

Brandl, J. E. (1980). Policy evaluation and the work of legislatures. In L. A. Braskamp & R. D. Brown (Eds.), Utilization of evaluative information. San Francisco: Jossey-Bass.

Braskamp, L.A., & Brown, R.D. (Eds.). (1980). Utilization of evaluative information. San Francisco: Jossey-Bass.

Braskamp, L.A., Brown, R.D., & Newman, D.L. (1978). Credibility of a local educational program evaluation report: Author source and client characteristics. American Educational Research Journal, 15, 441-450.

Brickell, H.M., Aslanian, C.B., & Spak, L.J. (1974). Data for

decisions: An analysis of evaluation data needed by decision-makers in educational programs. New York: Policy Studies in Education.

Brooks, C.R. (1984, April). Administrative decision-making and program evaluation use. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans. (ERIC Document Reproduction Service No. ED 252 554)

Brown, E.K. (1973, April). Changes in school management's needs for evaluation information. Paper presented at the annual meeting of the American Educational Research Association, New Orleans. (ERIC Document Reproduction Service No. ED 074 094)

Brown, R.D. (1980). The evaluator as artist: Primitive, impressionistic, or paint-by-number? CEDR Quarterly, 13, 4-6.

Brown, R.D., & Braskamp, L.A. (1980). Summary: Common themes and a checklist. In L.A. Braskamp & R.D. Brown (Eds.), Utilization of evaluative information. San Francisco: Jossey-Bass.

Brown, R.D., & Newman, D.L. (1982). An investigation of the effect of different data presentation formats and order of arguments in a simulated adversary evaluation. Educational Evaluation and Policy Analysis, 4(2), 197-203.

Brown, R.D., Newman, D.L., & Rivers, L. (1980). Perceived need for evaluation as an influence on evaluator's impact on decision-making. Educational Evaluation and Policy Analysis, 2, 67-73.

Brown, R.D., Newman, D.L., & Rivers, L. (1985). An exploratory study of contextual factors as influences on school board evaluation information needs for decisionmaking. Educational Evaluation and Policy Analysis, 7(4), 437-445.

Burry, J. (1983). Federal requirements affecting local evaluation use. Los Angeles: Center for the Study of Evaluation. (ERIC Document Reproduction Service No. ED 238 934)

Burry, J. (1985). Documenting evaluation use: Guided evaluation decisionmaking. Los Angeles: Center for the Study of Evaluation. (ERIC Document Reproduction Service No. ED 266 154)

Caplan, N. (1980). What do we know about knowledge utilization? In L.A. Braskamp & R.D. Brown (Eds.), Utilization of evaluative information. San Francisco: Jossey-Bass.

Carver, R.P. (1993). The case against statistical significance testing, revisited. Journal of Experimental Education, 61(4), 287-292.

Cohen, D.K. (1972). Politics and research: Evaluation of social action programs in education. In C.H. Weiss (Ed.), Evaluating social action programs: Readings in social action and education. Boston: Allyn & Bacon.

Cook, T.D. (1974). Sesame Street and the medical and tailored models of summative evaluation research. In J.G. Abert & M. Kamrass (Eds.), Social experiments and social program evaluation. Cambridge: Ballinger.

Cook, T.D. (1978). Utilization, knowledge-building, and institutionalization: Three criteria by which evaluation research can be evaluated. In T.D. Cook et al. (Eds.), Evaluation studies review annual (Vol. 3). Beverly Hills: Sage

Publications.
Cousins, J.B., & Leithwood, K.A. (1986). Current empirical research on evaluation utilization. Review of Educational Research, 56(3), 331-364.
Cronbach, L. J., Ambron, S.R., Dornbusch, S.M., Hess, R.D., Hornik, R.C., Phillips, D.C., Walker, D.F., & Weiner, S.S. (1980). Toward reform of program evaluation. San Francisco: Jossey-Bass.
Daillak, R.H., Alkin, M.C., & White, P. (1978, April). Studies of school evaluation practice. Paper presented at the annual meeting of the American Educational Research Association, Toronto. (ERIC Document Reproduction Service No. ED 170 345)
Datta, L. (1978, April). The new Congressionally mandated studies: Notes toward species identification. Paper presented at the annual meeting of the American Educational Research Association, Toronto. (ERIC Document Reproduction Service No. ED 156 850)
Datta, L. (1979). O thou that bringest the tidings to lions: Reporting the findings of educational evaluations. Paper presented at the annual Johns Hopkins University National Symposium on Educational Research, Baltimore.
Davis, H.R., & Salasin, S.E. (1975). The utilization of evaluation. In E.L. Struening & M. Guttentag (Eds.), Handbook of evaluation research (Vol. 1). Beverly Hills: Sage.
Dawson, J.A., & D'Amico, J.J. (1985). Involving program staff in evaluation studies: A strategy for increasing information use and enriching the data base. Evaluation Review, 9(2), 173-188.
Deal, T.E., & Rallis, S.F. (1980). Promoting interaction among producers and users of educational knowledge. In Interorganizational arrangements for collaborative efforts. Portland, OR: Northwest Regional Educational Laboratory.
Dickey, B. (1979, October). Effective problem-solving for evaluation utilization. Paper presented at the annual meeting of the Evaluation Research Society, Minneapolis.
Duggan, J.G., Talmage, H., & Rasher, S.P. (1983, April). Client use of evaluation findings: An examination of salient variables. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal. (ERIC Document Reproduction Service No. ED 228 297)
Eason, S., & Thompson, B. (1988, October). The effect of persuasion on the perception of program evaluation information. Paper presented at the annual meeting of the American Evaluation Association, New Orleans. (ERIC Document Reproduction Service No. ED 301 596)
Eisner, E.W. (1980, April). Forms of representation in educational evaluation. Paper presented at the annual meeting of the American Educational Research Association, Boston.
Fletcher, J.L. (1972, April). A model for data-based school improvement. Paper presented at the annual meeting of the American Educational Research Association, Chicago. (ERIC Document Reproduction Service No. ED 065 586)
Fullan, M. (1979, May). Conceptualizing problems of curriculum

implementation. Paper presented for the Symposium on Curriculum Inquiry in Canada, Victoria.

Fullan, M. (1980). An R&D prospectus for educational reform. In Interorganizational arrangements for collaborative efforts. Portland, OR: Northwest Regional Educational Laboratory.

Goldberg, M. (1978, April). New York City program administrators' perceptions of the helpfulness of outside mandated evaluations on administrative decision-making in ESEA programs. Paper presented at the annual meeting of the American Educational Research Association, Toronto. (ERIC Document Reproduction Service No. ED 156 720)

Goodrich, T.J. (1978). Strategies for dealing with the issue of subjectivity in evaluation. Evaluation Quarterly, 2, 631-645.

Gorham, W. (1970). Getting into action. Policy Sciences, 1, 169-176.

Granville, A.C. (1977, April). An experimental study of factors that influence school decisions: Implications for evaluation. Paper presented at the annual meeting of the American Educational Research Association, New York. (ERIC Document Reproduction Service No. ED 139 824)

Granville, A.C. (1978, April). The impact of the project developmental continuity evaluation. In A.C. Granville, J.M. Love, R.D. Matz, L.J. Schweinhart, & A.G. Smith (Eds.), The impact of evaluation: Lessons drawn from the evaluations of five early childhood education programs. Proceedings of a symposium presented at the annual meeting of the American Educational Research Association, Toronto. (ERIC Document Reproduction Service No. ED 166 212)

Greene, J.C. (1988a). Communication of results and utilization in participatory program evaluation. Evaluation and Program Planning, 11, 341-351.

Greene, J.C. (1988b). Stakeholder participation and utilization in program evaluation. Evaluation Review, 12(2), 91-116.

Guba, E.G. (1969, April). An overview of the evaluation problem. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles. (ERIC Document Reproduction Service No. ED 031 814)

Guba, E., & Lincoln, Y. (1981). Effective evaluation. San Francisco: Jossey-Bass.

Guba, E.G., & Stufflebeam, D.L. (1970). Evaluation; The process of stimulating, aiding, and abetting insightful action. Bloomington: Indiana University. (ERIC Document Reproduction Service No. ED 055 733)

Gurel, L. (1975). The human side of evaluating human service programs: Problems and prospects. In M. Guttentag & E.L. Struening (Eds.), Handbook of evaluation research (Vol. 2). Beverly Hills: Sage.

Guskin, A.E. (1980). Knowledge utilization and power in university decision making. In L. A. Braskamp & R.D. Brown (Eds.), Utilization of evaluative information. San Francisco: Jossey-Bass.

Guttentag, M., & Struening, E.L. (1975). The handbook: Its purpose and organization. In M. Guttentag & E.L. Struening (Eds.), Handbook of evaluation research (Vol. 2). Beverly Hills: Sage.

Haenn, J.F. (1980, April). Reasons why evaluations and testing don't inform. Paper presented at the annual meeting of the American Educational Research Association, Boston. (ERIC Document Reproduction Service No. ED 187 733)

Hall, G., & Loucks, S. (1977). A developmental model for determining whether the treatment is actually implemented. American Educational Research Journal, 14, 263-276.

Haller, E. (1974). Cost analysis for educational program evaluation. In J. J. Popham (Ed.), Evaluation in Education. Berkeley: McCutchan.

Hamilton, G.E. (1980, April). Title I evaluation from the perspective of the local school. Paper presented at the annual meeting of the American Educational Research Association, Boston.

Hansen, J.B. (in press). [Review of The program evaluation standards: How to assess evaluations of educational programs]. Educational and Psychological Measurement.

Havelock, R.G. (1968). Dissemination and translation roles. In T.L. Eidell & J.M. Kitchel (Eds.), Knowledge Production and Utilization in Educational Administration (Columbus, Ohio) and Center for the Advanced Study of Educational Administration (Eugene, Oregon) Career Development Seminar, October, 1967. Eugene, OR: University of Oregon Press.

Hayman, J. (1979, April). The case for internal linkage: Functional integration within the organization. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Herman, J.L. (1990). Political and practical issues in improving school boards' use of evaluation data. Los Angeles: Center for the Study of Evaluation. (ERIC Document Reproduction Service No. ED 324 792)

Holley, F. (1979, April). Catch a falling star: Promoting the utilization of research and evaluation findings. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

House, E.R. (1973). School evaluation: The politics and process. Berkeley: McCutchan.

Huberman, M. (1990). Linkage between researchers and practitioners: A qualitative study. American Educational Research Journal, 27(2), 363-391.

Ingison, L.J. (1979, October). Some realities and constraints regarding the use and usefulness of policy-relevant evaluations. Paper presented at the annual meeting of the Evaluation Research Society, Minneapolis.

Isaac, S. (1980, April). Increasing the relevance of national evaluation to schools: A school district's viewpoint. Paper presented at the annual meeting of the American Educational Research Association, Boston.

Johnson, M.D. (1978, March). Evaluation as a model for decision-oriented research. Paper presented at the Association for Institutional Research Annual Forum. (ERIC Document Reproduction Service No. ED 154 743)

Johnston, J. (1978, April). What we know about the researcher and research methods in evaluation. Paper presented at the annual

meeting of the American Educational Research Association, Toronto.

Joint Committee on Standards for Educational Evaluation. (1981). Standards for evaluations of educational programs, projects, and materials. New York: McGraw-Hill.

Joint Committee on Standards for Educational Evaluation. (1994). The program evaluation standards: How to assess evaluations of educational programs (2nd ed.). Newbury Park, CA: SAGE.

Kelezo, G.J. (1974). Program evaluation: Emerging issues of possible legislative concern relating to the conduct and use of evaluation in the Congress and executive branch. Washington, D.C.: Congressional Research Service, Library of Congress.

Kilbourne, R., & DeGracie, J. (1979, April). The use of L.E.A. research at the local level: The picture of a dropout. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

King, J.A. (1988). Research on evaluation use and its implications for evaluation research and practice. Studies in Educational Evaluation, 14(3), 285-299.

King, J.A., & Thompson, B. (1981). Evaluation utilization: An annotated bibliography. New Orleans: Orleans Parish School Board. (ERIC Document Reproduction Service No. ED 204 363)

King, J.A., & Thompson, B. (1983a). How principals, superintendents view program evaluation. NASSP Bulletin, 67, 46-52.

King, J.A., & Thompson, B. (1983b). Research on school use of program evaluation: A literature review and research agenda. Studies in Educational Evaluation, 9, 5-21.

King, J.A., Thompson, B., & Pechman, E. (1981). Evaluation utilization: A bibliography. New Orleans: Orleans Parish School Board. (ERIC Document Reproduction Service No. ED 207 984)

Knorr, K.D. (1977). Policymakers' use of social science knowledge: Symbolic or instrumental? In C.H. Weiss (Ed.), Using social research in public policymaking. Lexington, MS: Heath.

Law, A. (1980). In M.C. Alkin & A. Law, A conversation on evaluation utilization. Educational Evaluation and Policy Analysis, 2, 73-79.

Lenihan, K.J. (1977). Telephones and raising bail: Some lessons in evaluation research. Evaluation Quarterly, 1, 569-586.

Leviton, L.C., & Boruch, R.F. (1983). Contributions of evaluation to education programs and policy. Evaluation Review, 7(5), 563-598.

Leviton, L.C., & Hughes, E.F.X. (1979). Utilization of evaluations: A review and synthesis. Evanston, IL: Center for Health Services and Policy Research, Northwestern University.

Leviton, L.C., & Hughes, E.F.X. (1981). Utilization of evaluations: A review and synthesis. Evaluation Review, 5(4), 525-548.

Lyon, C.D., Doscher, L., McGranahan, P., & Williams, R. (1978). Evaluation and school districts. Los Angeles: Center for the Study of Evaluation.

McClintock, C. (1985, April). Clarifying information needs for program evaluation. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago. (ERIC

Document Reproduction Service No. ED 257 880)
McColskey, W.H., Altschuld, J.W., & Lawton, R.W. (1985). Predictors of principals' reliance on formal and informal sources of information. Educational Evaluation and Policy Analysis, 7(4), 427-436.
McLean, L.D. (1979). The chastening of educational research. Ontario: Ontario Institute for Studies in Education.
Mathis, W. (1980, April). The use of state assessment data as a response and generator of educational policy. Paper presented at the annual meeting of the American Educational Research Association, Boston.
Matuszek, P., & Holley, F. (1977, April). Evaluation for decision-making: A case study. Paper presented at the annual meeting of the American Educational Research Association, New York. (ERIC Document Reproduction Service No. ED 137 402)
Meltsner, A.J. (1976). Policy analysts in the bureaucracy. Berkeley: University of California Press.
Mitchell, D.E. (1980, April). Social science impact on legislative decision-making: Process and substance. Paper presented at the annual meeting of the American Educational Research Association, Boston. (ERIC Document Reproduction Service No. ED 186 347)
Mitchell, J. (1990). Policy evaluation for policy communities: Confronting the utilization problem. Evaluation Practice, 11(2), 109-114.
Newman, D., & Brown, R.D. (1980, April). An exploratory investigation of decision-making style and locus of control as factors related to choice of school evaluation strategies. Paper presented at the annual meeting of the American Educational Research Association, Boston.
Newman, D.L., Brown, R.D., & Rivers, L.S. (1983). Locus of control and evaluation use: Does sense of control affect information needs and decision making? Studies in Educational Evaluation, 9(1), 77-88.
Newman, D.L., Brown, R.D., & Rivers, L.S. (1987). Factors influencing the decision-making process: An examination of the effect of contextual variables. Studies in Educational Evaluation, 13(2), 199-209.
Newman, D.L., Brown, R.D., Rivers, L.S., & Glock, R.F. (1983). School boards' and administrators' use of evaluation information: Influencing factors. Evaluation Review, 7(1), 110-125.
Newman, D.L., Bull, K.S., Brown, R.D., & Rivers, L.S. (1986). Locus of control as an influence of school evaluation needs. Evaluation Review, 10(4), 536-552.
Novak, C.D. (1977, April). An involvement approach to the evaluation of local district programs. Paper presented at the annual meeting of the American Educational Research Association, New York. (ERIC Document Reproduction Service No. ED 152 819)
Oman, R.C., & Chitwood, S.R. (1984). Management evaluation studies: Factors affecting the acceptance of recommendations. Evaluation Review, 8(3), 283-305.
Orlandi, L.R., & Conslave, R.W. (1977). Blaming the victim of

evaluation. CEDR Quarterly, 10, 3; 7.
Patton, M.Q. (1978). Utilization focused evaluation. Beverly Hills: Sage.
Patton, M.Q. (1988). The evaluator's responsibility for utilization. Evaluation Practice, 9(2), 5-24.
Patton, M.Q. (in press). [Review of The program evaluation standards: How to assess evaluations of educational programs]. Educational and Psychological Measurement.
Pflum, G.D., & Brown, R.D. (1984). The effects of conflict, quality, and time on small group information use and behavior in evaluative decision making situations. Evaluation and Program Planning: An International Journal, 7(1), 35-43.
Polivka, L., & Steg, E. (1978). Program evaluation and policy development: Bridging the gap. Evaluation Quarterly, 2, 696-707.
Raizen, S.A. (1978, April). Introductory comments. In A.C. Granville, J.M. Love, R.D. Matz, L.J. Schweinhart, & A.G. Smith (Eds.), The impact of evaluation: Lessons drawn from the evaluations of five early childhood education programs. Proceedings of a symposium presented at the annual meeting of the American Educational Research Association, Toronto. (ERIC Document Reproduction Service No. ED 166 212)
Randall, R.S. (1969, April). Knowledge about decision processes and information. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles. (ERIC Document Reproduction Service No. ED 031 814)
Revicki, D., & Rubin, R. (1980, April). Models for measuring program implementation: A review and critique. Paper presented at the annual meeting of the American Educational Research Association, Boston.
Ripley, W.K.. (1985). Medium of presentation: Does it make a difference in the reception of evaluation information? Educational Evaluation and Policy Analysis, 7(4), 417-425.
Rippey, R.M. (1973). Studies in transactional evaluation. Berkeley: McCutchan.
Ross, J.A. (1980). Decision rules in program evaluation. Evaluation Review, 4, 59-74.
Rossi, P.H. (1972). Booby traps and pitfalls in the evaluation of social action programs. In C.H. Weiss (Ed.), Evaluating action programs: Readings in social action and education. Boston: Allyn & Bacon.
Rutman, L. (1977). Barriers to the utilization of evaluation research. Paper presented at the annual meeting of the Society for the Study of Social Problems, Chicago.
Saks, M.J. (1980). The utilization of evaluation research in litigation. In L.A. Braskamp & R.D. Brown (Eds.), Utilization of evaluative information. San Francisco: Jossey-Bass.
Sanders, J.R. (1994, April). The process of developing national standards that meet ANSI guidelines. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.
Shapiro, E. (1973). Educational evaluation: Rethinking the criteria of competence. School Review, 81, 523-549.
Shapiro, J. (1979, April). Assessing the impact of evaluation: A

univariate time series technique. Paper presented at the annual meeting of the American Educational Research Association, San Francisco. (ERIC Document Reproduction Service No. ED 174 689)

Simon, H.A. (1957). Administrative behavior. New York: Macmillan.

Stake, R. (1973). Evaluation design, instrumentation, data collection, and analysis of data. In B.R. Worthen & J.R. Sanders (Eds.), Educational evaluation: Theory and practice. Worthington, OH: Charles A. Jones.

Steele, S.M. (1973). Contemporary approaches to program evaluation. Washington, D.C.: Education Resources Division of Capitol Publications.

Stevens, W.F., & Tornatzky, L.G. (1980). The dissemination of evaluation; an experiment. Evaluation Review, 4, 339-354.

Stevenson, J.F. (1979, October). Evaluation utilization in human service organizations: Is there a role for the in-house evaluator in assisting utilization? Paper presented at the annual meeting of the Evaluation Research Society, Minneapolis.

Stufflebeam, D.L. (1994, April). Old and more recent history of the inception and revision of the program evaluation standards. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Stufflebeam, D.I., Foley, W.J., Gephart, W.J., Guba, E.G., Hammond, R.L., Merriman, H.D., & Provus, M.M. (1971). Educational evaluation and decision making. Itasca, IL: Peacock.

Suchman, A.E. (1972). Action for what? A critique of evaluation research. In C.H. Weiss (Ed.), Evaluating action programs: Readings in social action and education. Boston: Allyn & Bacon.

Thompson, B. (1980). Validity of an evaluator typology. Educational Evaluation and Policy Analysis, 2, 59-65.

Thompson, B. (1981, October). Communication theory as a framework for evaluation use research: Evaluation as persuasion. Paper presented at the annual meeting of the Evaluation Network Society, Austin. (ERIC Document Reproduction Service No. ED 209 330)

Thompson, B. (Ed.). (1982, April). Evaluation issues from an LEA perspective. Symposium presented at the annual meeting of the American Educational Research Association, New York. (ERIC Document Reproduction Service No. ED 216 046)

Thompson, B. (1989, March). Five steps for improving evaluation studies by using different data analysis methods. Paper presented at the annual meeting of the American Educational Research Association, San Francisco. (ERIC Document Reproduction Service No. ED 306 244)

Thompson, B. (1993). The use of statistical significance tests in research: Bootstrap and other alternatives. Journal of Experimental Education, 61(4), 361-377.

Thompson, B., & King, J.A. (1981a). A critique of evaluation-use research methods. CEDR Quarterly, 14, 19-21. (Reprinted in Louisiana Education Research Journal, 1982, 7, 1-6)

Thompson, B., & King, J.A. (1981b, April). Evaluation utilization: A literature review and research agenda. Paper presented at

the annual meeting of the American Educational Research Association, Los Angeles. (ERIC Document Reproduction Service No. ED 199 271)

Tittle, C.K. (1977). Evaluation and decision making: Developing a method to link program funding decisions and outcome evaluation. Paper presented at the annual meeting of the American Educational Research Association, Toronto.

Turner, S.D., Hartman, J., Neilsen, L.A., & Lombana, J. (1988). Fostering utilization through multiple data gathering methods. Studies in Educational Evaluation, 14(1), 113-133.

Van den Berg, G., & Hoeben, W.Th.J.G.. (1984). The practice of evaluation research and the use of evaluation results. Studies in Educational Evaluation, 10(3), 309-321.

Weiner, S.S., Rubin, D., & Sachse, T. (1977). Pathology in institutional structures for evaluation and a possible cure. Stanford, CA: Stanford Evaluation Consortium.

Weiss, C.H. (1972). Utilization of evaluation: Toward comparative study. In C.H. Weiss (Ed.), Evaluating action programs: Readings in social action and education. Boston: Allyn & Bacon.

Weiss, C.H. (Ed.). (1977). Using social research in public policy making. Lexington, MS: Lexington Books.

Weiss, C.H. (1979, October). Conceptual issues in measuring the utilization of research and evaluation. Paper presented at the annual meeting of the Evaluation Research Society, Minneapolis.

Weiss, C.H. (1987, October). Evaluations for decisions: Is anybody there? Does anybody care?. Paper presented at the annual meeting of the American Evaluation Association, Boston.

Wholey, J.S., Scanlen, J., Duffy, H., Fukumoto, J., & Vogt, L. (1970). Federal evaluation policy: Analyzing the effects of public programs. Washington, DC: The Urban Institute.

Wilensky, H.L., & Lebeaux, C.N. (1958). Industrial society and social welfare. New York: Russell Sage.

Williams, W., & Elmore, R.F. (1976). Social program implementation. New York: Academic Press.

Williams, W., & Evans, J.W. (1969). The politics of evaluation: The case of Head Start. Annals of the American Academy of Political and Social Science, 385, 118-132.

Wise, R.I. (1978, April). What we know about the decision-maker and decision settings. Paper presented at the annual meeting of the American Educational Research Association, Toronto.

Wise, R.I. (1980). The evaluator as educator. In L. A. Braskamp & R.D. Brown (Eds.), Utilization of evaluative information. San Francisco: Jossey-Bass.

Worthen, B.R., & Sanders, J.R. (1973). Educational evaluation: Theory and practice. Worthington, OH: Charles A. Jones.

Young, C.J., & Comptois, J. (1979). Increasing Congressional utilization of evaluation. In F. Zweig (Ed.), Evaluation in legislation. Beverly Hills: Sage.

Zepeda, R.A. (1980, April). What the models cannot tell us: Information needs for program planning. Paper presented at the annual meeting of the American Educational Research Association, Boston.